Part B lite (Just the Basics) Review Checklist for Protocols and SOPs

Roy Irwin, NPS, WRD

April 12, 2005 Draft

Please send peer review comments to roy_irwin@nps.gov

Table of Contents:

Introduction:

Note: The following is condensed from the much longer version of Part B at http://science.nature.nps.gov/im/monitor/protocols/water qualityPartB.doc).

As suggested in generic VS guidance (K.L. Oakley, L.P. Thomas, and S.G. Fancy, 2003. Wildlife Society Bullet 31(4), reprint at http://science.nature.nps.gov/im/monitor/protocols/ProtocolGuidelines.doc, all sampling protocols will include three basic sections:

A. Protocol Narrative
B. Protocol Standard Operating Procedures (SOPs), and
C. Protocol Supplementary Materials.

So one item on any protocol checklist is whether or not the protocol follows the organization above, is complete, and has a table of contents that helps one determine where things are. Either the protocol narrative or a separate SOP should include a discussion of who will do the monitoring and who will train them and how often (recurrent training and is Quality Assurance/QA basic). Is there a SOP that clearly defines protocol variables and how to measure them?

The following text summarizes the basics of what has to be in water quality and other aquatic protocol SOPs [usually including a quality assurance/quality control (QA/QC) or a quality assurance project plan (QAPP) SOP] to meet checklist ("Checklist for Review of Vital Signs Monitoring Plans," on the Internet at http://science.nature.nps.gov/im/monitor/docs/MonitoringPlanChecklist.doc, hereafter referred to as "the checklist") review requirements.

This summary can also be used for the basics that should be included in Phase 1, 2, and 3 monitoring plan chapters. In most cases, the planning process is iterative, with very general statements in the plan chapters becoming more detailed in the subsequent protocols and SOPs.

Among the basics that need to be covered in the narrative and SOPs are (adapted from USGS wildlife monitoring website at http://testweb-pwrc.er.usgs.gov/monmanual/):

WHAT are you going to measure?
WHERE are you going to put your sampling points?
HOW are you going to measure it?
WHEN (HOW FREQUENTLY) are you going to measure it?

Some have found Part b to be too long, so Part b lite is an effort to provide networks with an extremely condensed version they can use for a last minute check to see if all the checklist basics have been covered in protocol narratives and attached SOPs.

The full version of part B can still be used for much more detailed discussions of the entire planning process and how all the pieces (review of past data, questions, target populations, representativeness, precision, detection limits, etc.) all fit together.

Summary of Information from Past Data

Checklist question: For water quality monitoring, has information content of available past aquatic data (for each waterbody being considered for monitoring) been adequately summarized in terms of hints of trends or other important issues of concern?

The emphasis should not be on who is monitoring where, but what does it all mean.

A table listing 303d waters should be included in the protocol narrative, along with a note that the most recent WRD Designated Use and Impairments database (at http://www1.nrintra.nps.gov/wrd/dui/) has been consulted and that any differences with Network versions of the 303d lists have been logically reconciled. The 303 documentation in the protocol narrative should include as much spatial detail (from where to where) as possible, typically more detail than related discussions in the Background section in chapter 1 of the central monitoring plan.

Objectives and Questions:

It is easiest to plan monitoring if general objectives are rephrased into more detailed questions in the protocol narrative. The monitoring is then planned in such a way that questions can be answered with the data collected. In Chapter 1 of the monitoring plan itself questions can be somewhat general, but the questions should be made more detailed in both time and space when they are addressed in more detail each protocol narrative. These detailed questions should make sense in comparison with summary discussions for representativeness and named target population(s) about which inferences will be made.

This part of the protocol narrative is where the following checklist item should be addressed "Does the protocol narrative identify specific measurable objectives such as thresholds or trigger points for management actions?" If this issue is not addressed here, there should be a "point to" hyperlink showing where it is addressed.

This part of the protocol narrative should also summarize which questions and/or sites were selected to ensure monitoring of a 303d impaired water body or a very pristine water body that the park wants to keep that way. WRD has suggested that at roughly 2/3 of the sites should be in one of those two categories (see Part A guidance). What monitoring will be done to help answer GPRA reporting goals?

Include Detailed SOPS for All Field and Lab Methods

Field Collection or Measurement SOPs should detail exactly what will be done in the field. Exactly how will field measures be done?

Lab SOPs should detail exactly how everything is done in the lab. If a standard State of USGS or EPA method is used, it should be written out or attached in its entirety. If the agency (EPA, USGS, etc.) changes the method, will the NPS also change in the same way? The SOPs should be detailed enough to allow third parties to reproduce the methods and to allow determinations of data comparability (see further discussions below).

Examples of details to be included in protocols rather than in the central monitoring plan:

More details on sampling locations and method specifics.

For example, in chapter 4 a table might say that Chlorophyll *a* was the parameter to be monitored. Chapter 5 would give the method details, Chlorophyll *a* is to be monitored using field water collection procedures of the USGS field manual and lab method x (USGS Schedule 1637 method, or EPA method 445.0, or APHA method 10200H-4, or whatever is selected).

If flow or water level is to recorded, will it be qualitative or quantitative?

What field instrumentation will be needed?

What pre and post season activities are required?

How will samples be collected and preserved, what containers will be used, and what minimum holding times will be used? Unless otherwise justified, use suggestions in 40 CFR Part 136 to 136.3 and appendices.

Attach a QA/QC SOP to Each Aquatic Protocol

A key quote in the checklist that "For water quality monitoring, there should be a quality control SOP associated with each protocol that adequately documents quality control (QC) objectives for measurement sensitivity (detection limits), measurement precision, measurement systematic error (bias as percent recovery), data completeness (including adequacy of planned sample sizes and statistical power), and (if applicable for lab measurements only) blank control."

Since the checklist item (just above) is so brief, more information on what the WRD will be looking for in a QC SOP is provided herein.

For aquatic projects, and any other project where quality assurance and quality control (QA/QC) is important (where is quality not important?), QA/QC details should ordinarily be included in a separate QA/QC SOP attached to each protocol. It could also be called a QC SOP or a Quality Assurance Project Plan (QAPP) SOP. Whatever it is called, it should include sections explaining how each the following will be controlled and documented for individual parameters or suites of parameters:

1) Representativeness, Target Population, And Completeness
2) Data Comparability (Internal/NPS And External/Other Regional Data)
3) Measurement Sensitivity (Usually MDL And PQL) Detection Limits
4) Measurement Precision As Reproducibility And/Or Repeatability
5) Measurement Systematic Error/Bias/Percent Recovery (Still Wrongly Called Accuracy By Some)
6) Blank Control Bias (Usually For Chemical Lab Work Only)

If additional detail on any of the topics listed above is located anywhere other than in the QA/QC SOP, a summary of what will be done to control each of the issues listed above should be included in the QA/QC SOP and it should be made clear exactly where the other detail is in the monitoring plan or protocol. For example, if representativeness and target populations are fully in explained in the protocol narrative or in the chapter on sampling design in the plan, then the representativeness section of the QA/QC SOP should clearly "point to" the section where the subject is fully covered. However, often the protocol narrative text will be more general, with the details related to individual sites and individual measurements/parameters (characteristic in STORET-speak) being more fully explained in the representativeness section of the QA/QC SOP.

To obtain data comparability, it is OK and even desirable to use well established QA/QC procedures of another federal agency (USGS, NOAA, EPA, NAWQA, or EMAP) or a State agency, but the source and the measurement quality objectives and SOP details for sensitivity/detection limits, precision, systematic error/bias, and blank control (for chemical lab work only) need to be listed in the QA/QC SOP because the other groups may change their SOPs as time goes along, and we need to have solid

documentation of what we started with and/or what will be used in our long term monitoring.

QC details such as measurement quality objectives will often be different for individual parameters/vital signs/characteristics to be measured. However, there may be cases where the same QC measurement quality objective might be given for several parameters in a suite of vital signs included in one protocol. For example, if a network decided to use EPA marine EMAP QC SOPs to obtain maximum data comparability with EPA and State marine EMAP data, they could specify a precision repeatability measurement quality objective of a relative percent difference (sample size 2) or a relative standard deviation (sample size 3 or more) of 10% for several parameters to be measured in the field, including include pH, temperature, DO, specific conductance, salinity depth, light transmittance (PAR), turbidity, and Secchi depth. However, in many other cases measurement quality objectives would be different for different parameters and could simply be listed in a QC SOP table included as part of each protocol. Thus a protocol for water column parameters measured in the field would typically have different measurement quality objectives than a protocol for nutrient parameters measured in the lab. However, in both cases, a table in a separate QC SOP in each protocol could list the measurement quality objectives for each applicable parameter.

Be careful with QA/QC terminology. Words and phrases such as representativeness, sensitivity, detection limits, accuracy, precision, repeatability, reproducibility, error, systematic error/bias, and uncertainty, have been used for different concepts by different groups. The confusion in water quality and contaminants QA/QC terminology has been so widespread that it brings to mind a “Tower of Babel” (everyone speaking different languages, no one understanding each other) scale of confusion. Some care has been taken herein (and in more detail in Part b) to explain the right terminology and to standardize on NIST and ISO terminology wherever possible.

If the QA/QC of another agency is not adopted, or if the QC details come from multiple sources or are brand new, before completing the QA/QC SOP, a final check should be made to make sure that the measurement process will be controlled in some documented and defendable manner. The networks need to document what will be done for each of the issues listed below (doing nothing is not an option), but the networks need not “go overboard.” However, at minimum, reviewers will be looking for common-sense documentation related to each of the following QA/QC basics:

Representativeness and Target Populations:

Given what is known about variability in time and space, how will the sampling scheme insure that the value obtained will be representative of the target population being studied (checklist, op cit.)? The target population is simply the larger universe of all possible values (bounded in time and space) that one is sampling from and wishes to make statistical inferences about**.**

There is growing recognition that unless care is taken to ensure representativeness, data can be of little value, no matter how good the measurement performance is for precision, bias, detection limits, etc. In other words, ensuring data

quality means not only insuring analytical quality but also sample representativeness (a QC basic) of the target population given the questions to be answered (see part B).

Like many other facets of VS planning, representativeness is typically assured in an iterative fashion. First Chapter 4 should briefly describe how representativeness will be assured. Later in the protocol section discussing study design, add more detail about how representativeness will be achieved by the study design selected. Finally, in the QA/QC SOP attached to each protocol, include any other needed details about how representativeness will be achieved for each parameter to be monitored, including what is known about true heterogeneity in time and space for each parameter to be monitored, average values, and variability in the various strata. It would be appropriate to mention in the representativeness section what statistics are planned and how those fit in with representativeness, the target population, and the questions to be answered.

Since representativeness is so related to study design, statistics, target populations, and questions to be answered some of the information related to representativeness will be in several different places. Therefore, liberal use of “point to” statements or hyperlinks is suggested so that readers can easily find the discussions and tables.

In the protocol narrative or QA/QC SOP, explain what the target population is in time and space detail. Explain how sampling will be done in a manner that insures the data collected will be representative of the target population and useful in answer identified questions. Justify the sampling design (random, spatially-balanced random with unequal (but not zero) chance of selection, other EMAP variants, biased design to sample historical bridge sites, etc.) and how it insures representativeness. Where in the process is randomness or stratified random sampling involved? Even if a historical bridge site is sampled, there should be some randomness in exactly where a stream is sampled near the bridge. Generic VS guidance calls for some randomness in the process somewhere.

Chapter 4 requirements for study design are well covered in the generic VS study plan guidance and the checklist (op cit.). Of particular interest in aquatic projects are the water monitoring-specific portions of the checklist. The checklist calls for a map of water bodies & sites to be monitored and an accompanying table of frequency of sampling events, and the number and different types of samples to be taken during each sampling event, all as part of chapter four.

How many samples will be taken and to achieve data completeness, what percentage can fail and not result in sample sizes too small to answer the identified questions? These issues are just as important for biological monitoring as for water chemistry monitoring.

In a few cases, networks might be able to justify (based on data comparability, difficult access and other factors) assuring representativeness in ways that USGS WRD has historically (sample cross sections and composite the sample, try to use integrator sites, check to see if well mixed first, etc.). However, if absolutely no randomness is involved, the rationale should be justified on a network-specific basis. If no site-specific randomness is involved, say in exactly where to sample at the site (or in a strata), why not? Why do the pros of the targeted design outweigh the cons, given the money available, the need to continue historical trend data, and USGS's findings that

representativeness is assured given documented studies (site them if available) in the region, etc.?

When sampling near bridges, whenever possible, it is preferable to sample far enough way from bridges to minimize bridge effects (deicer salts, dust, vehicle pollutants, trash, and changed hydrology etc.) to help with representativeness with this general area of the stream and not just the (often unusual) conditions right at the bridge.

A good example of probabilistic sample designs that allow inference to a broader target population than single targeted site is the GRTS design of the HTLN network. Another good solution and variant on spatially balanced probability design is the Systematic, Unequal Probability Sampling Design utilized by the Northern Colorado Plateau Network (NCPN). Some networks use hybrid sampling plans that include both 1) targeted sites (such control sites or historically sampled bridge sites) to answer site-specific or other limited inference questions and 2) probability-selected sites that allow for broader inferences to larger areas of the park or waterbody. A good discussion of tradeoffs between targeted designs, probabilistic designs, and hybrid designs in water quality monitoring is EPA's discussion at http://www.epa.gov/maia/html/llstr4.htm.

The SFAN freshwater quality protocol is a good example of reiterating the basics of study design with more detail in the protocol narrative. Whereas maps in chapter 4 of the central monitoring plan might show where water bodies are in the network, maps in the protocols should show location details of each sampling site, reach, (or area or strata where random selection will occur) in greater detail. What types/kinds of sampling will be done (what is involved spatially and vertically by waterbody)? Will vertical profiling (e.g. non-flowing water bodies/lakes) or stream cross sections, or continuous monitoring be used? Where will randomness be included in the process? Will grab sampling or compositing be done? Will sites be rotated through on some yearly or multi-year basis? Some details could go into tables in the protocol narrative. Will water aquatic samples be co-located with other sampling or observations (habitat, fish, invertebrates, etc.)?

If some of the detailed explanations are in the plan itself and the decision is made that they need not be repeated in detail in the protocol, briefly recap them in the QC SOP section of representativeness and target populations, and provide a hyperlink or other "point to" in the QC SOP so the reader will know where to find the additional (more detailed) discussions.

Completeness, Sample Sizes, and Statistics

Data completeness goals are typically given as percentages in tables in the QA/QC SOP or QAPP and are developed by first estimating required sample sizes. Required sample sizes are usually driven by variability, and the probability of being able to detect a change magnitude of concern (or power for standard null hypothesis testing only, see section IV-C of part b) and required detection probabilities for threshold values or effect sizes of concern (see more below). Sample sizes need to be large enough to detect changes or effects of concern. The following question relates not only to completeness and required sample sizes and sampling design, but also to data analysis and statistics.

What is the probability of detecting an effect size of concern given the monitoring design and variability of the parameters and the statistics chosen? Generic Vital Signs guidance has stated that "Statistical detection limits, given typical sample variability and chosen sample sizes, shall be low enough to insure that such threshold values or trigger points can be detected" (Outline for Vital Signs Monitoring Plans, 2003). A later version of this was simplified but gets at the same issue and indirectly requires a similar thought process: "List the specific, measurable objectives for each vital sign selected for monitoring, and wherever possible, give the threshold value or "trigger point" at which some action will be taken" (Outline for Vital Signs Monitoring Plans, 2004, http://science.nature.nps.gov/im/monitor/docs/monplan.doc).

Since the answer the this question relates to so many other issues, we suggest networks put the answer in the data analysis SOP, but also include brief recaps or "point to" links from other related sections (the discussions of representativeness and completeness in the QA/QC SOP, and the sampling design discussions (Chapter 4 in the central monitoring plan).

An initial start at projecting sample sizes vs. the need to detect certain effect sizes might made by using simple sample size or power calculators to help estimate initial sample sizes needed. One often has to increase (10% is usually plenty) sample sizes calculated with parametric calculators due to lack of normally distributed data before using two sample nonparametric tests (see Part b for more detail).

Once an initial estimate of needed sample sizes are completed, one can then use the McBride detection probability calculator for equivalence tests (http://www.niwa.co.nz/services/statistical/detection/ to estimate the probability of detection of various effect-sizes given the sample size chosen. Say one has decided to use only 10 samples to save money. The detection probability calculator shows that a small sample size of 10 has:

1. Less than a 15% probability of detecting a 20% (of the mean) effect size if a null hypothesis t test is used.
2. Less than a 70% probability of detecting a proof of safety, based on an inequivalence test and the threshold assumption that a 20% effect size is within a bioequivalent (de minimis effects) range and that therefore that much change is a 'safe' change to be expected with normal variation.

The reason this calculator does not require variance as an input variable is that normal distribution detection probabilities are based on a given proportion of the true standard deviation and the (dimensionless) effect size is the independent parameter. Power curves (and operating characteristic curves) have the effect size as a percentage of the mean on their x axis and the detection probability on the y axis.

Power is only a big concern when using a standard null hypothesis test (usually not a great idea anyway), so calculating detection probabilities using equivalence or inequivalence tests is usually a better start.

However, these matters are typically complex, so the calculators above should be only a first (reality check on sample sizes needed, get you in the general ballpark) step

and before finalizing sampling designs, needed sample sizes and data completeness goals, we suggest a professional applied statistician be consulted to consider the following:

> Knowledgeable applied statisticians can help networks estimate needed sample sizes using modeling approaches. Standard power analyses are legitimate only in a small set of cases where analyses allow for a simple power analysis and statistical hypothesis testing is deemed reasonable. More often, analyses are too complicated to do a plug-in power calculation and multiple hypotheses are to be considered simultaneously. This requires Monte Carlo simulation-based approach to determining sample size and power.
> **(http://science.nature.nps.gov/im/monitor/meetings/Austin_05/PLukacs_SampleSize.doc).**

Once required sample sizes are estimated, one then estimates of the number of projected samples that are not likely to produce usable data, recognizing that it is very rare that 100% of all planned samples are successfully obtained and also pass all QC data acceptance criteria. Extra samples are then planned and completeness goals are usually some percentage (such as 80-90%) of the total number of samples planned.

Data Comparability (Internal/NPS and External/other regional data)

For Internal Data Comparability: What will be done to maximize temporal and methodological consistency in NPS data? Control typically involves limiting changes in internal NPS methods or timing of sampling to help insure our own newer data is comparable to our older data. When methods are changed, bias from the change should be documented in the data analysis SOP (see section below entitled "Archive Cumulative Bias from Method Changes in the Data Analysis SOP" for more detail).

For External Data Comparability: What will be done to achieve comparability with other regional data sets (USGS, States, NOAA, CERCLA, etc)? Again, what will be done to insure our NPS data are comparable enough to the data from other state and federal agencies that need to be convinced our data is credible and comparable, given our purposes for monitoring? Is our NPS data comparable enough to other important outside data sets that the two sets of data could be combined for purposes of determining trends or making management or regulatory decisions? Has the chemical lab being used passed federal round robin blind sample checks or been approved the NELAP or the applicable State regulatory water quality agency?

A final check should be made to make sure both the lab and the field method SOPs attached to the protocol are detailed enough to allow for reproducibility of exactly the same methods by third parties. Are they also detailed enough to allow judgments about the comparability of the data with the data of other agencies? Perfectly comparable data can be merged and analyzed together without introducing problems.

These issues are just as important for biological monitoring as for water chemistry monitoring. Interagency efforts are now being made to come up with acceptance criteria for comparability, something now rarely utilized.

Why Document Quality Control?

We are now moving from more qualitative subjects to documenting quantitative measurement quality objectives for quality control (QC) data quality indicators. Before doing so, we will first answer the following question:

> Why do we need to quantitatively control and document QC data quality indicators like measurement sensitivity, measurement precision, measurement systematic error/bias, and blank control?

The short answer is that most states and regulatory agencies require us to do so as part of a required quality assurance project plan (QAPP) and/or due to State credible data laws. It is also required by NPS WRD and the VS checklist.

If QC measurement quality objectives are not controlled and documented, most regulatory agencies will not consider our data to be credible enough to use in Clean Water Act decisions or processes (like section 303d use impairment determinations and Total Maximum Daily Loads/TMDLs). It is expensive to collect data, so the extent possible if should be useful for multiple purposes, especially purposes that will help resource protection or management, or be helpful related to GPRA goals.

Even if no outside entity is was "making us" do so, there are logical reasons why we should control and document the performance of QC data quality indicators:

1. The scientific community has known since the 1930's that measurement processes need to be quantitatively controlled to produce credible data, so not doing so in 2005 and beyond would be difficult to justify.
2. Limiting problems with measurement precision and measurement bias to specified amounts controls how imperfect the measurement process can be and still be and produce acceptable results. Without such controls, we have no way to estimate how badly the measurement process is performing. Measurement uncertainty could be extremely large and we wouldn't know it. In any case, we would not be able to document how great the measurement uncertainty was.
3. Accordingly, if ever challenged in court, not controlling and documenting measurement performance for QC indicators (sensitivity, precision, bias, etc.) would be impossible to defend.
4. This is long term monitoring, and controlling and documenting data quality indicators gives our data a better chance to be considered credible in future years. In future years we don't want our data thrown out because someone has decided that all data without detection limits or other quality control documentation are unacceptable. Judging data without QC documentation as not credible or useful is already happening and the tendency to do so can be expected to become more common.
5. In long term monitoring, method, SOP, and staff changes are inevitable and documenting changes in measurement performance is therefore even more important than for short term projects. Documenting QC performance of the old and new methods helps one determine whether or not a change in values was the result of a true change in the environment,

or whether it is simply the result of a SOP, meter, or staff change that changed measurement system performance.

6. NPS WRD requires that all VS or WRD funded water quality data are archived in Modernized STORET. STORET contains fields for detection limits, precision, systematic error/bias as % recovery, blank control results, etc.
7. NPSTORET is even more insistent on having detection limits. In NPSTORET, if one enters a result with detection conditions 'Not Detected', 'Present, below Quantification Limit', or 'Present, above Quantification Limit', one will not be able to enter a value. Data with these conditions requires entry of the relevant limit.
8. We need some QC results for common-sense and data interpretation reasons. For example, we need measurement sensitivity results to understand whether or not the analyte is present and how big of a change in the concentration of the analyte is believable as a real change, rather than a random error in the measurement process.
9. We logically agree with the DOI information quality guidelines that emphasize QA/QC basics such as ensuring and maximizing the “quality, utility, objectivity, and integrity of the information” and to ensure a high degree of transparency about data and methods used. Optimal transparency would include admitting that is data are not perfect but that the degree of imperfection could not have exceeded MQOs specifications for precision, bias, and sensitivity.
10. Reproducibility is not only a QC basic, it is a “sound science” basic. Unless one documents measurement performance characteristics, it will be very difficult for another party to reproduce the result independently.

Measurement Sensitivity

When measuring water quality chemicals at very low levels, a system that can accurately measure and detect a very low concentration is more sensitive than one that can only detect the presence of the analyte at higher concentrations. The more sensitive the measuring system, the lower the low-level detection limits are.

Toxic chemicals can be hazardous at very low levels, and there is typically a concern about whether or not they are present in Parks, even at very low levels. Likewise, some pristine waters in the NPS have very low concentrations of nutrients, and the Parks want to keep them that way. Both of these scenarios lend themselves documenting and controlling measurement sensitivity with the lowest practicable low-level detection limits. Low level detection limits have been the most common way sensitivity has been handled in the past, and for many water quality parameters, they are still critical.

However, even if we are always measuring in higher measurement ranges (well above the low-level quantitation detection limits); we still need to control measurement sensitivity. For some parameters measured in the field (pH, temperature, conductivity, biological observations, physical habitat observations, etc.), one seldom if ever encounters extremely low levels. In these higher measurement ranges, the smaller the (true) change that a measuring system can (reliably and accurately) detect, the more

sensitive it is. For these types of measurements or observations, low level detection limits are less relevant and/or less helpful, and an alternative or additional way to estimate measurement sensitivity is suggested below.

Low Level Detection Limits (MDLs and PQLs)

Minimum Requirement: In the QA/QC SOP, list (a table is fine) pre-project targets (and how often they will be estimated once monitoring begins) for the following low-level detection limits:

1. A **semi-quantitative** method detection limit (MDL) and,
2. A **quantitative** Practical Quantification Limit (PQL).

For NPS standardization, the MDL and PQL are the suggested defaults. The MDL and PQL are to be calculated as follows:

MDL:

List the target standard EPA method detection limit (MDL), semi-quantitative detection limit for each parameter to be measured in the QA/QC SOP. Labs should be instructed to calculate the MDL as explained in Appendix B to 40 CFR Part 136—Definition and Procedure for the Determination of the Method Detection Limit—Revision 1.11 (http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=fadf72a44c330e5dd957e61dddab97e7&rgn=div9&view=text&node=40:21.0.1.1.1.0.1.6.2&idno=40**).**

Although most labs and even some EPA staff and published EPA methods do not always use all the steps suggested by EPA to calculate an MDL, most at least eventually use the central equation of Method Detection Limit (MDL) = t times S, where, t = the one-sided Student's t-value for seven replicate (precision repeatability) samples. In this equation, for 7 replicate samples, t = 3.143, so MDL = 3.143 times the sample standard deviation for the 7 replicate measurements of a blank. The same equation is used in the Standard Methods Book definition of an MDL. According to EPA, the MDL is the lowest value we really believe with 99% confidence is different than zero (a one sided comparison, which is why we use the one-sided t value). The same equation is usually used to estimate estimated detection limits (EDLs), usually with fewer steps than estimating an MDL, and often with low level standards and not blanks.

However, to avoid confusion, alternative semi-quantitative detection limits (EDL, LOD, IDL, LLD, etc., see part B for details) should ordinarily not be listed. The exceptions include: 1) If a USGS lab is used, the USGS alternative to the MDL (the LT-MDL) can be listed along with how it is calculated and how often, or 2) If there is no good way to calculate or find an MDL but an estimated detection limit can be found (sometimes the case for bacteria or chlorophyll) and calculated, the EDL can be defined and used..

PQL:

Unless otherwise justified, define and calculate the practical quantitation limit (PQL) as 3.18 times the MDL. The resulting value is the same as the (low level) lower quantification limit (LQL, a STORET-specific term), but it is suggested that the more universal PQL terminology be used rather than LQL except when dealing with STORET.

If a network needs to do so because it is dealing with multiple labs or cannot achieve quantitative detection limits at 3.18 times the MDL, the network could alternatively define the PQL as 5 (rather than 3.18) times the MDL. If this is done, a rationale should be provided to justify doing so.

However, there are disadvantages for using 5, and it should be a justified the exception rather than a default choice. Using 5 would result in being able to report fewer low level values (see following section). Also, some EPA methods specify the use of 3.18 and top experts in the field now consider 3.18 to be sufficiently high to protect against false negatives, to be the value most commonly used, and to have other advantages over 5 (for details see Part B, or D. Helsel. 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. Wiley. 288 pp., http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471671738.html, or Environmental Protection Agency. 2003. Technical Support Document for the Assessment of Detection and Quantitation Concepts, EPA publication no. EPA-821-R-03-005, available at **http://www.epa.gov/waterscience/methods/det/dqch1-3.pdf).**

To avoid confusion and for NPS standardization, use the PQL rather than alternative quantitative detection limits phrases or acronyms. There are so many acronyms for these alternatives, that it is easy to become confused in the alphabet soup: Alternatives include minimum levels (MLs), minimum reporting levels (MRLs), reporting levels (RLs), limits of quantitation (LOQs), minimum quantitation limits (MQLs), sample quantitation limits (SQLs), Contract-Required Quantitation Limits (CRQLs), or an interlaboratory quantitation estimate (IQE). Instead of using these terms, convert all such quantitative limits to PQLs. Those desiring to understand more about these terms can read part B.

There are two exceptions where terms other than the PQL can be used:

1. When dealing with STORET, the phrase lower quantitation limits (LQL) can be considered a synonym for a PQL.
2. If a USGS lab is used, the USGS alternative to the PQL (The LRL) may be used instead of the PQL. The LRL is defined as two times the USGS LT–MDL (http://water.usgs.gov/owq/OFR_99-193/level.html).If the network is going to use USGS laboratory reporting levels (LRLs), explain how often they will be calculated and reported.

No matter what quantitative detection limits are used, how they are calculated should be explained in the sensitivity/detection limit part of the QA/QC SOP. Once monitoring begins and new data is put in STORET, it should also be put in STORET metadata, as explained in the next section.

For toxic chemicals, it is particularly important that the lab can achieve PQL quantitative detection limits that are below the benchmark, water quality standard or criteria, or other threshold levels known to be associated with harmful effects.

How Will Values Below The PQL or MDL Be Reported And Analyzed?

The measurement sensitivity/detection limit section of the QA/QC SOP should explain how data below any of the listed detection limits will be handled, not only for reporting into data bases, but also for data analyses. An acceptable option (and one already adopted by some networks) is to state that the recommendations in the recent Helsel Book (D. Helsel. 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. Wiley. 288 pp.**,** http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471671738.html**)** will be followed.

NPS data needs to go into STORET, and the Helsel book (op cit.) considers the modernized STORET default recommendation for writing to a database to be fully acceptable, so we are adopting this as a default NPS recommendation. Modernized STORET suggests that we not report into a database any value higher than the MDL but lower than the PQL. Instead, the detection condition field is set to "Present, below Quantification Limit". With that detection condition, STORET automatically enters "*Present <QL" in the result field. A major advantage of this approach is that no "estimates" are treated as quantitative when in fact they are not quantitative.

In later statistical analyses values between the MDL and PQL are best interpreted using either an interval-censored method (parametric), or a rank-based method (nonparametric) where all in-between values are represented as the same tied rank. The older recommendation of censoring to half the MDL is clearly no longer recommended. Helsel also gives recommendations for how not to report into data bases (for example, never report single values below the MDL or even the PQL, and do not report nondetects as half the detection limit. One should also not report nondetects as a negative ("-") sign followed by the actual MDL value, because someone invariably decides it really is a negative number (Helsel, op cit.).

Values above the PQL are classified in EPA's modernized STORET database with the detection condition of "Detected and Quantified" This is ideal, and according to EPA STORET Staff, this is optimally the only choice which permits reporting a single number.

Although not recommended in the Helsel book (op cit), for the special case of NPS analyses of "precautionary principle" comparisons with standards or criteria, one might choose to censor all data below the PQL to the exact value of the PQL, but that is only a very special (worst-case, trying to totally avoid false negatives) short-term case of data analysis, and one would never substitute the value of the PQL in a long term storage data base field for measured concentrations.

Alternative Measurement Sensitivity For Measuring Above the PQL (AMS)

In a sense, the MDL and PQL are commonly used strategies for bounding and thereby controlling measurement uncertainty and measurement sensitivity when one is trying to measure extremely low values, at or near the lowest value that can reliably be measured (the PQL), or the lowest value that can be distinguished from zero (the MDL).

When one is consistently measuring above the PQL in the normal quantitative measurement range, one can use standard National Institute of Standards and Technology (NIST) methods and terminology.

NIST works with its international counterpart, the International Organization for Standardization (ISO) to standardize measurement methods and terminology used in science and engineering worldwide. They have reached consensus on measurement basics. These include:

1. No measurement is perfect. Each is an approximation, and
2. Individual measurement data points are not complete unless accompanied by a statement about the uncertainty of that approximation.

NIST has acknowledged that standard NIST/ISO methods to calculate NIST "expanded measurement uncertainty" are not applicable for very low (below quantitative-PQL detection limits) ranges of measurement (N. Taylor and C. E. Kuyatt. 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results NIST Publication TN 1297, http://physics.nist.gov/Document/tn1297.pdf).

However, if one is consistently measuring values well above the very difficult low-level ranges, MDLs and PQLs are not very relevant or very helpful in dealing with measurement sensitivity. In some cases (pH and temperature, for example), no blank or other zero-value solutions are available. In other cases, zero-value solutions would not be relevant to the measurement ranges of interest.

When measuring conductivity in the field, or when making biological or physical observations in the field, a low level detection limit is usually not very helpful. However, in all of these cases where standard MDLs and PQLs are difficult or not relevant, controlling measurement sensitivity is a still a QC basic that should not be ignored.

Since we are not considering the lower end of the range in which one cannot use NIST/ISO standard calculations for expanded measurement uncertainty, there is no reason not to do things the standard way. The APHA (Standard Methods Book, http://standardmethods.org/) also recommends that expanded measurement uncertainty be calculated the same way as NIST and ISO, lending even more credibility and standardization to the method.

Therefore, the NPS default suggestion is to calculate an Alternative Measurement Sensitivity (AMS) based on NIST expanded precision uncertainty using a sample size of 7 and 99% probability. This satisfies two needs at the same time: 1) the need to control measurement sensitivity when in the normal quantitative measurement range, and 2) the need to have a plus or minus value to put in the STORET "analytical procedure description" text box.

To calculate AMS in a way that is as functionally analogous to sensitivity based on MDLs and PQLs as possible, one should use the same factors used in MDL and PQL calculations: 1) The same number (7) of replicate measurements of one sample, and 2) the same (99%) level of confidence.

It may seem like the discussion has been shifted from sensitivity to precision. However, measurement sensitivity is usually derived from looking at precision results in an especially rigorous way but doing so less often than one samples normal precision QC samples. That is what we are suggesting here.

To determine AMS sensitivity, measure one typical-concentration sample 7 times and use a 99% level of confidence. For contrast, when one controls precision separately (and more often), one measures a single QC precision sample twice (not 7 times) and one doesn't worry about 99% confidence, as explained in the next chapter.

The terminology AMS is used to emphasize that what we are dealing with here is measurement sensitivity, not just standard QC precision samples, and not MDLs or PQLs.

AMS results are not recorded in detection limit fields. Instead they should be recorded in the STORET metadata "analytical procedure description" text box. AMS should be calculated in the same manner that NIST/ISO/APHA "measurement precision expanded uncertainty" is calculated for a 7 replicate samples and a 99% level of confidence (N. Taylor and C. E. Kuyatt. 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results NIST Publication TN 1297 (http://physics.nist.gov/Document/tn1297.pdf).

A plain language step-by-step for doing this is provided as follows: first, one samples one NORMAL (not a blank as was the case for MDLs or PQLs) environmental sample seven times. Those 7 results are used to calculate the sample standard deviation. Next the sample standard deviation times is multiplied by 3.708 (the 99% confidence middle t value for sample size 7) to get expanded precision uncertainty = AMS.

The result of this calculation in some respects functionally analogous to a PQL but uses the two-sided (middle) t value because we are not only interested in one side (distinguishing a value from zero) but in the two-sided issue of how large of a difference between two individual values we can justify as actually being a true difference (from one another) and not due to measurement process noise.

The AMS (as expanded precision uncertainty) should be calculated and reported no less than at least once every sampling season, and more often until a reasonably consistent range is developed. It should be re-calculated when something significant (for example, the person or the measuring instrument) changes and may have changed measurement sensitivity.

The result should be called "alternative measurement sensitivity based on expanded precision uncertainty" (AMS for short) to distinguish it from standard low-level MDLs and PQL estimates of sensitivity.

The AMS result may be used not only to fill in blanks in STORET but also to bound measurement uncertainty on each single data point. This allows us to be consistent with modern and transparently honest scientific thought that no measurement or observation is perfect, each is an estimate. Just as confidence intervals express the uncertainty about a mean, there is an interval of uncertainty around each single measurement data point.

If measurement precision and sensitivity were very good, a result for a single data point might be something like 45.676 plus or minus 0.003. In the more common (for field monitoring) scenario where neither sensitivity nor precision are that good, the result for a single point might be reported as something more like 50 plus or minus 30.

Either way, the result can be entered into STORET in the plus or minus field for "precision" in the CHEMICAL DATA RESULT ENTRY BOX. Bounding uncertainty in this way is a more modern alternative to using rounding rules to decide how many

significant figures one should carry in final result (after all multiplications and divisions or other manipulations have been completed).

Unlike the MDL or PQL, data is typically not censored based on AMS/measurement precision expanded uncertainty values. However, as a statistical analysis strategy, one could take the worst case end of the range. For example, if the highest pH value considered safe was 9.0, and the single value measured was 8.9 plus or minus 0.3, one might say the single value could be as high as 9.2 and therefore may exceed the criteria.

Many biological inventory and monitoring projects have not historically estimated measurement sensitivity. However, there is usually no reason why one person could not calculate AMS after measuring one sample 7 times (or perhaps have one sample measured by 7 different biotechs, or something similar). It may take some ingenuity in difficult cases, or (in cases of destructive sampling) may require a sampling nearby samples (see "precision plus" discussion in precision chapter below). However, with careful thought, it should usually be possible to develop a common-sense way to adapt the AMS/expanded measurement precision uncertainty logic into AMS functional analogs for various types of biological monitoring. The key is to try do so in a way that "makes sense" while still addressing the issue of logically estimating and controlling measurement sensitivity and uncertainty.

Can one list both low level detection limits and alternative measurement sensitivity for field measurements? Yes, as explained above, there are separate places in STORET to put both results. In fact, in most cases, for field-measurements it would be optimal to do both.

Is it ok to use the lower end of the applicable measurement range (in manufacturer's specifications) as an estimate of the MDL and for reporting to STORET? This is not ideal and should never be done if the lower end of the specification range is zero. Doing this would not be recommended where true MDLs are needed (for toxics and very low level nutrients in very pristine lakes). In these cases, proper MDLs and PQLs should be calculated.

However, if one is always measuring values well above the lower end of the range and does not encounter really bad precision (the field instrument will not settle down on one reading but just keeps changing, or the precision RPD is above precision MQOs when measuring values greater than 2 x an MDL or a roughly estimated MDL), one should not worry about continual reporting MDLs and PQLs for field measurements.

MDLs and PQLs are not required in STORET and are only required in NPSTORET if one reports the detection conditions of either 'Not Detected' or 'Present, below Quantification Limit.' So if one never encounters this scenario, one need not enter a MDL or PQL into NPSTORET or STORET should simply calculate and report a more appropriate type of sensitivity (AMS as explained above).

What does one put in the MDL and PQL tables in the QC SOP in this situation? In many cases one can use another program's defaults (a state or EMAP for example). One might also just enter a code in the table that is explained in more detail at the end of the table. The code explanation might explain that as long as precision MQOs are met above twice the MDL and the field readings settle down to one value, these other-agency or less stringently-estimated MDLs and PQLs will be considered sufficient. NEMI (www.nemi.gov) sometimes gives the lower end of the calibrated range as a 'range-

derived' lower detection limit, and this or some other common-sense rough estimate of a MDL might be used in QC tables when bad precision does not require a more stringent standard low level MDL. In these cases, AMS will be periodically calculated and reported.

If one encounters low levels where good precision is impossible, a situation that might occur in some pristine environments, normal MDLs and PQLs should be calculated and reported.

Resolution

Measurement resolution relates to a single measurement value and is usually somehow related to the fineness of the measurement scale, but beyond that, little seems standardized. It is typically not acceptable to use the resolution specifications of the manufacturer of a field meter for either a detection limit or for precision, since there is no consistency between manufacturers in how resolution is estimated.

Resolution is a term used by makers of field measurement meters but (for good reason) the word resolution is not typically seen in environmental quality assurance project plans (QAPPs). Resolution specifications may have little to do with precision or sensitivity in the field, but may have been developed for competitive advantage in ideal lab situations.

For most water quality or contaminants applications, use of the word resolution is not justified, since those using the word resolution are often talking about other more commonly understood QC concepts, such as precision, uncertainty in accuracy, sensitivity, detection limits (a special case of sensitivity when signals are low), or measurement uncertainty. Since these concepts are defined in detail separately herein and more universally understood, there is typically no need to address "resolution" separately in water quality or contaminants QAPPs or QA/QC SOPs.

Certain GIS/Remote sensing and non-linear biological categorization applications may be exceptions. In remote sensing, the word resolution is often used for a concept more broadly recognized in other disciplines as sensitivity. In any case, if the word resolution is used, how it is determined should be defined in detail.

Measurement Precision as Reproducibility and/or Repeatability

Precision (actually imprecision, but according to tradition and common practice most ignore that) is the variability of repeated independent measures of the same object**.** ISO defines precision as "The closeness of agreement between independent test results obtained under prescribed stipulated conditions." Contributors to lack of perfect precision include random measurement error and random sampling error. Precision is the result of random (up and down) errors and (unlike systematic error/bias) does not relate to the true value or the specified value.

NIST clarifies that the "prescribed stipulated conditions" should include documentation about whether the precision is "precision under repeatability conditions" or "precision under reproducibility conditions" (B. N. Taylor and C. E. Kuyatt. 1994. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results NIST Publication TN 1297 (http://physics.nist.gov/Document/tn1297.pdf

In more plain language, precision is simply the variability of the different observations compared to each other. For sample size two (a duplicate), in water quality work, precision is most commonly expressed as a relative percent difference (RPD). For larger sample sizes it is most often expressed as a relative standard deviation (RSD or coefficient of variation expressed as a percent). As explained in the section above, it can also be expressed as multiple of a sample standard deviation, though this is typically not done in water quality or contaminants work except when one is calculating detection limits or expanded measurement uncertainty.

Precision is not only estimated in a rigorous way every now and then to estimate sensitivity/AMS/measurement uncertainty (as discussed above), it is also estimated much more often in a much less rigorous way (sample size 2 rather than 7, using typical concentration environmental samples rather than blanks) to monitor measurement precision. This is a standard QC requirement and is the subject of this section. Precision duplicates are typically done often enough to document that the measurement process is remaining "in control." When controlling precision in batches or groups, QC precision samples are usually collected every 20 samples or so and then measured twice (duplicates) to control measurement precision on a regular basis.

A check list of items to be included in a separate QC SOP for each protocol is provided as follows. For each measurement done in the field or lab, are the following adequately covered?

- A measurement quality objective (MQO) for precision.
- Will the MQO be used as a data acceptance performance standard? (It usually should be)
- What is the data comparability source of the MQO (State, USGS, EPA-EMAP, RCRA, CERCLA, CWA, etc.)?
- Is precision being controlled in the context of repeatability, reproducibility, reproducibility plus, or some combination (specify)?
- How will precision be calculated and reported?
- Will the raw number results be reported in addition to summary statistics like RPDs, RSDs, sample standard deviations, or multiples of standard deviations? (The raw numbers should also be reported).
- How often will precision be estimated and reported?

Standard precision MQOs can often be summarized in a table, sometimes along with systematic error, method detection limits, and blank control MQOs (See SFAN freshwater quality protocol example).

Precision in context of repeatability is the scenario where nothing in the measurement process changes. For chemical lab measurements, repeatability MQOs are typically used, but if multiple labs, multiple instruments, or multiple staff become factors, it becomes precision in the context of reproducibility.

Precision in the context of reproducibility is typical of long term monitoring, since there will typically be changes in staff and instruments, or sometimes different staff and different instruments are even used in the same network during one season. So often precision is controlled as reproducibility.

Precision "Reproducibility Plus" is our NPS terminology for field duplicates when two samples that are not exactly the same are taken in close proximity in time and/or space. Since the samples may not be identical, the "plus" part of the phrase is a tip off that an additional potential source of variability is present. In this case, two potential sources of variability are being lumped, lack of prefect measurement precision plus potential true sample heterogeneity. When taking this approach, it should typically be done in addition to (rather than instead of) more conventional estimates of measurement precision. The text or tables should explain how often both precision and precision plus will be estimated to determine how much of the variability is due to random error in the measurement process vs. how much is due to true sample heterogeneity. If only one type of precision is to be estimated and controlled, unless otherwise justified, it should not be precision reproducibility plus.

Regardless of the type of precision controlled, usually duplicate samples every 10-20 samples or every sampling batch, or every field sampling day (specify which) are used as precision QC samples.

The text should document the extent to which precision MQOs will be used as data rejection criteria (unless otherwise justified, they should be data rejection criteria). So if the MQO for a particular parameter is that a relative percent difference cannot exceed a plus or minus 30%, if the samples exceed that value, all values associated with that batch or that QC sample should be discarded, and recalibration or other adjustments should be done until the MQO can be met.

Many types of biological inventory and monitoring projects have not historically estimated measurement precision. However, there is growing recognition of the need to do so, and one can usually find a common-sense way to control and document measurement precision in biological projects. Often one can simply measure something twice to get a duplicate answer (see Part B for more detail).

Measurement Systematic Error/Bias/Percent Recovery
(Still wrongly called accuracy by some)

Systematic error/bias is the systematic or persistent distortion of a measurement process that causes errors in one direction (high or low). On the measurement scale of concern, systematic error and bias are usually considered synonyms. Though it has commonly been used in this context in the past, the word accuracy should not be used for the concept of bias since uncertainty in accuracy can only be estimated after factoring in not only systematic error/bias, but also precision.

In water quality and contaminants work, systematic error/bias is usually expressed as a as % recovery (with the correct or expected answer being considered 100%) of an interval such as 80-120%. That particular example could be shortened to 20% as long as it is made clear that 20% means a percent recovery interval of $\pm$ 20% (same as 80-120% recovery). The raw values used to calculate these percentages should also be reported to allow one to later look at the results in other ways (such as long term or multi-lab means or standard deviations).

For each measurement done in the field or lab, are the following adequately covered?

- A systematic error/bias measurement quality objective (MQO), such as % recovery must be within 80-120%.
- Will the MQO be used as a data acceptance performance standard?
- What is the data comparability source of the MQO (State, USGS, EPA-EMAP, RCRA, CERCLA, CWA, etc.)?
- How will systematic error/bias be calculated and reported?
- How often will systematic error/bias be estimated and reported?

Unless otherwise justified, MQOs should be rejection criteria. So if the MQO for a particular parameter is that a % recovery cannot be worse that 70-130%, then if the recovery 60% or 140%, all values associated with that batch should be discarded, and recalibration or other adjustments should be done until the MQO can be met.

If one value (say water color) is being measured to estimate another value (say chlorophyll a), how will bias and accuracy (including a precision component) be controlled and estimated? Will average observed to expected ratios be used, and how? Will root mean square error techniques be used, and exactly how and how will the result compare to detection limits that use multiples of the standard deviation?

Many types of biological inventory and monitoring projects have not historically estimated measurement systematic error/bias. However, there is growing recognition of the need to do so, and one can usually find a common-sense way to control and document measurement bias in biological projects. One strategy sometimes used is to consider a senior expert's answer right or expected (100%) and a rookie trainee's answer wrong (see Part B for recommended strategies).

Blank Control Bias (usually applicable to chemical lab work only)

Blank QC samples (samples known to be free of the analyte being measured) are tested to see if they have been contaminated with the analyte during collection, handling, and processing steps. Contamination results in a positive bias in the concentration. Unless otherwise justified, for lab chemical measurements of toxic chemicals, metals, pesticides, or nutrients, MQOs for blank control shall be listed in the QA/QC SOP. Unless otherwise justified, blank samples above the qualitative method detection limit (MDL).

For each chemical measurement done in the lab are the following adequately covered?

1. A blank control measurement quality objective (MQO), if applicable.
2. What types of blanks will be controlled (trip blanks, lab blanks, etc.)
3. Will the MQO be used as a data acceptance performance standard?
4. Will data reported be adjusted by adding concentrations found in blanks? If not, how will blank control be accomplished (reduce contamination and re-run the samples?).
5. What is the data comparability source of the MQO (State, USGS, EPA-EMAP, RCRA, CERCLA, CWA, etc.)?
6. How will blank control be calculated and reported?
7. How often will blank control be estimated and reported?

Biological inventory and monitoring projects have not historically done blank control. However, if the scenario of wrongly assigning a number value when the true value is zero seems likely, it might be possible to develop a common-sense way to control bias from blanks.

Include Calibration Details

Instrument calibration details should be included either in the QC SOP or in a separate calibration SOP (checklist, op cit.). If these details are somewhere else (perhaps in the field or lab method SOPs), there should be a "point to" in the QC SOP so that the reader will be able to find them.

Include a Data Analysis SOP

Are there recommendations for routine data summaries and statistical analysis to detect change? How often will reporting and trend analyses be done? Does this SOP or the protocol narrative describe the frequency of testing and review of protocol effectiveness?

The data analysis SOP should include a discussion of the data analyses (including statistics) planned, who will do them, how often, and ensure that adequate staff time and project funding is set aside for this very important task. Most of the proposed statistics should be worked out with a statistician before protocols are completed.

Some networks have also appropriately said that statistics will usually be handled according to the recommendations of the following two text books:

> Helsel, D.R. and R.M. Hirsch 1992. Statistical Methods in Water Resources. Studies in Environmental Science 49, Elsevier Publishing, NY, this on is on the net for free at http://water.usgs.gov/pubs/twri/twri4a3/pdf/twri4a3.pdf.
>
> D. Helsel. 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. Wiley (op cit., see detection limit section).

Archive Cumulative Bias from Method Changes in the Data Analysis SOP

Method, equipment, and personnel changes are inevitable in long term monitoring. The requirement of overlapping old and new measurement methods is in Oakley et al (op cit.), but is often overlooked. How long will the old and new methods be overlapped to determine changes in measurement bias or precision? It is suggested that single (identical) samples be measured by both the old and new methods (or by old and new personnel) at least 7 times (whenever possible), with the results of the average positive or negative bias (as well as any changes in precision as reproducibility expanded uncertainty) archived in a place that future data users can easily find it. It is suggested that the cumulative results of the bias over the years be detailed in the Data Analysis SOP in each protocol, with "point to" hyperlinks from other places people might look, such as the protocol revision log, each field and lab SOP for methods, the data management SOP, the data management section or data acquisition parts of the central monitoring plan, and

the precision and bias discussions in the QC SOP. Even small changes in measurement bias can accumulate and become significant over time.

The goal would be for someone 100 years later to be able to discover the effect of the various changes in measurement bias. If a future user could look in the data analysis SOP and discover that 90 years ago there was a method change that resulted +2% change (on identical samples) from the previous method, then 80 years ago there was another method change that resulted in another change of +4%, then later a plus 3%, then later a -1%, they would be better equipped to separate true trends from the cumulative effects of numerous measurement method changes. This issue is important enough that some redundancy provided by the multiple "point to" links from other places seems prudent.

Include STORET Details in a Data Management SOP

Documentation and planning in the SOP needs to include matching the network's characteristics/parameters with the official standardized EPA list of 337,378 (as of 1/5/2005) characteristics (found in tblDef_TSRCHAR in NPSTORET (for questions, contact Dean_Tucker@NPS.GOV) or at http://nrdata.nps.gov/Programs/Water/storetcharacteristics/storetcharacteristics.zip).

End of Part B lite. More detail on each of the topics in Part B lite is found in the long version of Part B.